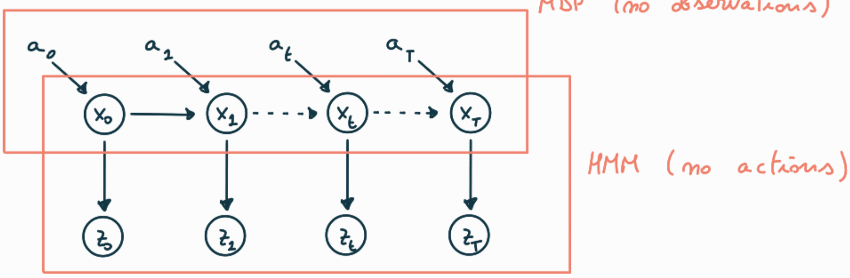
Partially Observable Markov Decision Process (POMDP)

Combines decision making of MDP and non-observability of HMM.

Partially observable : we cannot observe the states but we can observe them through the observations.
Decision process : we can choose the evolution of the system.



POMDP = MDP + HMM

the agent can decide the action and these actions will influence the evolution of the system

POMDP = $< X, A, Z, \delta, r, o >$

X : set of states
A : set of actions
Z : set of observations
$P(x_0)$ = probability distribution of the initial state
$\delta(x, a, x') = P(x' \mid x, a)$ probability distribution over transitions
$r(x, a)$: reward function
$o(x', a, z') = P(z' \mid x', a)$ probability distribution over observations

The solution of a POMDP is a policy, but we don't know the states
We can introduce a belief state : from the history of observations we can make an estimate of the probability distribution over the states and then use this estimate to choose the next action.

$$\pi : b_t \longrightarrow a_t$$

belief at time t

The set of all the possible beliefs is exponential with respect to the set of all the possible states.

Belief $b(x)$ = probability distribution over states

POMDP can be described as an MDP in the belief states, but belief states are infinite.

$$\left.\begin{array}{ll} B & = \text{set of belief states} \\ A & = \text{set of actions} \\ T(b, a, b') = & \text{probability distribution over transitions} \\ \rho(b, a, b') = & \text{reward function} \end{array}\right\} \text{MDP}$$

transform POMDP in a MDP and then solve the MDP with the known methods

policy: $\pi : B \longrightarrow A$

the problem is that with the transformation we move to a space exponentially bigger: this is not a practical approach.
There are ways of representing the belief space in a more compact way (approximation of the real transformation) that make the POMDP algorithms effective; in these cases we have no guarantee to get the optimal policy.

POMDP example: tiger problem